Visualization System to Analyze Browsing Trends of Internet Video Advertisements

Rika Miura
Ochanomizu University
Tokyo, Japan
g1820536@is.ocha.ac.jp

Hayato Ohya
Septeni Japan, Inc.
Tokyo, Japan
hayato.oya@septeni.co.jp

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
itot@is.ocha.ac.jp

***Abstract*—The market size of Internet advertising continues to grow. The creation and delivery of effective Internet advertising will become increasingly important in business. On the other hand, it is difficult to achieve an advertising effect if the target setting is inappropriate because Internet advertising is targeted and delivered to a specific group. Therefore, this study aims to discover effective Internet advertising target settings by evaluating Internet video advertising data with abandonment rate, click through rate and conversion rates and by developing a visualization system that takes into account combinations of advertising attributes. As a concrete example, we introduce the results of visualizing 224,355 anonymized records of advertisements distributed by Yahoo! JAPAN Ads and 89,400 anonymized records of advertisements distributed by LINE Ads.**

***Index Terms*—Visualization method, video advertising, abandonment rate, click through rate, conversion rate**

I. Introduction

Advertisements on the Internet continue to grow their market size year after year. According to the report by Dentsu Inc. [1], in 2021, Internet advertising expenditures in Japan reached 2.7 trillion Japanese yen and exceeded 2.4 trillion Japanese yen of the total advertising expenditures of the four mass media, i.e., newspapers, magazines, radio and television. In particular, video advertising grew significantly to 512.8 billion Japanese yen, 132.8% up from last year, and surpassed 500 billion Japanese yen for the first time. Due to this background, the creation and delivery of effective Internet advertising will become increasingly important in business.

Matsumoto et al. [9] analyzed marketing data by using clustering and principal component analysis to identify consumer characteristics. They clarified what kind of personality and consumption value characteristics consumers are influenced by advertisements. The results suggested that effective advertising methods differ from advertisement to advertisement. Zeng et al. [12] discovered the characteristics of advertisements disliked by users and preferred by users with cluster analysis. The above studies proved the necessity of selection of the appropriate delivery methods for each advertisement and creating advertisements that are preferred by consumers in order to promote effective advertising activities. In addition, Internet advertisements are delivered to specific groups unlike television and other media, and therefore, the expected effect cannot be achieved unless the appropriate target is set for each advertisement before delivery.

Besides, the target setting of Internet advertising is not always an alternative choice such as "female or male" but can be logical combinations of multiple attributes such as "female and 30s". It is often a bothering task to find out the trends of Internet advertising with combinations of attributes. In other words, it is still an open problem to visualize and analyze advertising data with combinations of advertising attributes. In this paper, we comprehensively evaluated video advertising with abandonment rate, click through rate, and conversion rate and developed the visualization system which takes into account the combination of advertisement attributes. This paper introduces our observation of the advertisement data by using this visualization system. The main contributions of this study are as follows.

1) Analysis of the advertisement data by evaluating comprehensively video advertisements with the three indicators below.
   - Abandonment rate (Percentage of video playback stopped in the middle)
   - Click through rate (Percentage of ad clicks)
   - Conversion rate (Percentage of app installations, product purchases, contract signings, and so on)
2) Development of a visualization system which takes into account the combination of advertising attributes for advertising data.

The remainder of this paper is organized as follows: Section 2 describes related research and Section 3 describes the proposed method. Section 4 discusses the results of the analysis. Section 5 summarizes this paper and discusses future issues.

II. Related Work

This section introduces existing studies for advertising data visualization and utilizing data for the purpose of improving advertising effectiveness.

A. Methods for visualizing advertising data

The video advertising data used in this study is multidimensional data. There have been various discussions on effective multidimensional data visualization methods. There have been several studies of visualization analysis specific to advertising data. Peng et al. [11] proposed an interactive analysis tool called TargetingVis. This tool enables advertising analysts to

2375-0138/23/$31.00 ©2023 IEEE
DOI 10.1109/IV60283.2023.00011

discover ideas to make advertising more effective by visualizing targeted advertisement delivery data and examining the delivery patterns of advertisers. TargetingVis uses the method for visualizing target combination patterns in the same way as in our study and suggests that this method is useful for discovering effective advertisement settings. Liu et al. [8] proposed MulUBA, a multi-view interactive system. This visualization system helps the analysis of online shopping behavior and attributes of customers by advertising analysts by combining multiple visualizations. These studies use click through rate and conversion rate as indicators of advertising effectiveness but do not use advertisement abandonment rate. On the contrary, our study aims to discover effective advertisement settings by comprehensively visualizing the browsing trends of video advertisements based on the three indicators of abandonment rate, click through rate, and conversion rate.

B. Utilization of Data for the Purpose of Improving Advertising Effectiveness

There have been many studies aimed at improving advertising effectiveness, in addition to the studies on visualization analysis methods as described in section 2.1. One of the representative researches which approach the improvement of advertising effectiveness is the study of click through rates prediction(CTR). Chen et al. [3] proposed an end-to-end integrated deep network to predict the CTR of advertisements. Zhou et al. [13] proposed the new model called DIEN for CTR prediction which takes into account the changing trend of the user interests. There exists also a lot of research not only CTR prediction but also analyzing advertisements effectiveness by using advertising data. Azimi et al. [2] analyzed the effect of advertisements appearance on users to assist creating high-quality creative. Kitada et al. [6] proposed a framework for correctly evaluating the effectiveness of advertising creatives by learning on unbalanced data using multi-task learning and a conditional attention mechanism. Ishikawa et al. [5] discussed the data which provide advertisers with clear quantitative indicators among ten types of basic advertisement data obtained from individual advertisements submitted to Yahoo! promotion advertisements. As a result, they found a combination of indicators that are significantly involved in conversion prediction. Zeng et al. [12] discovered what kind of advertisements make consumers uncomfortable by cluster analysis. They revealed that the clusters with the good advertisements were the ones that were simple, well-designed, and relevant to the interests of customers. Meanwhile, they found that the clusters which attracted the bad advertisements had list articles and political ones. Minjung et al. [10] analyzed differences in online behavior by consumer characteristics with clustering methods. Cheung et al. [4] investigated sizes and types of advertisements which might cause users to leave the advertisements. As mentioned above, studies on effective advertisement production support have been conducted from various perspectives; on the other hand, this study aims to improve advertisement effectiveness from the approach of visualization analysis.

III. Processing Flow of Observation and Visualization of Annotation Tasks

A. Dataset

In this study, we used anonymized datasets in the same format as the video advertisement data actually delivered on Yahoo! JAPAN Ads and LINE Ads.

B. Data Preprocessing

We created feature values for abandonment rate, click through rate, and conversion rate. This section describes each indicator in detail.

1) Video ad withdrawal rate: We evaluate advertisements by using the abandonment rate of video advertisements in this study.We calculated the six variables R_1 - R_6 listed below by using the four variables in Table 1 to generate clusters with similar trends in video ad abandonment and analyze the data. The six variables R_1 - R_6 represent the video ad abandonment rates between each time point.

TABLE I
Four values used to calculate the abandonment rate.

V_{p25}	Number of times the video was viewed up to 25% of the video play time
V_{p50}	Number of times the video was viewed up to 50% of the video play time
V_{p75}	Number of times the video was viewed up to 75% of the video play time
V_{p95}	Number of times the video was viewed up to 95% of the video play time

$$R_1 = \frac{V_{p25}-V_{p50}}{V_{p25}}, R_2 = \frac{V_{p25}-V_{p75}}{V_{p25}}, R_3 = \frac{V_{p25}-V_{p95}}{V_{p25}},$$
$$R_4 = \frac{V_{p50}-V_{p75}}{V_{p25}}, R_5 = \frac{V_{p50}-V_{p95}}{V_{p25}}, R_6 = \frac{V_{p75}-V_{p95}}{V_{p25}} \quad (1)$$

Click through rate / Conversion rate: The factors which cause users to leave video advertisements can be divided into positive and negative factors. Negative abandonment refers to the fact that the advertisement does not have an advertising effect and the user simply leaves the video advertisement. On the other hand, positive abandonment occurs when the advertisements were clicked. It is difficult to distinguish whether the abandonment is a positive or negative one if we evaluate advertisements based only on abandonment rates. In general, Internet advertisements are often evaluated by the click through rate (CTR) which is the number of clicks relative to the number of advertisements displayed, and the conversion rate (CVR) which is the rate at which the purpose of advertisement delivery is achieved after an advertisement is clicked. Therefore, we considered that it would be possible to distinguish between positive and negative abandonment factors by combining the abandonment rate, click through rate, and conversion rate of video advertisements. This study comprehensively evaluates advertisements by using the abandonment rate, click through rate, and conversion rate of video advertisements.

$$CTR = clicks/impressions \times 100 \quad (2)$$
$$CVR = conversions/clicks \times 100 \quad (3)$$

C. Dimensionality reduction and clustering

The six variables R_1 to R_6 related to the abandonment rate of video advertisements calculated in section 3.2.1 were mapped from a six-dimensional space to a two-dimensional space by applying a dimensionality reduction method. In addition, by applying a clustering process to the six variables R_1 to R_6, advertisements with similar abandonment tendencies belong to the same cluster. In this study, we observe what characteristics are observed among the advertisements which have similar trends in video ad abandonment rates. Principal component analysis was used as the dimensionality reduction method and the k-means method was used as the clustering method.

D. Visualization

We applied the following two visualizations for multidimensional training data.

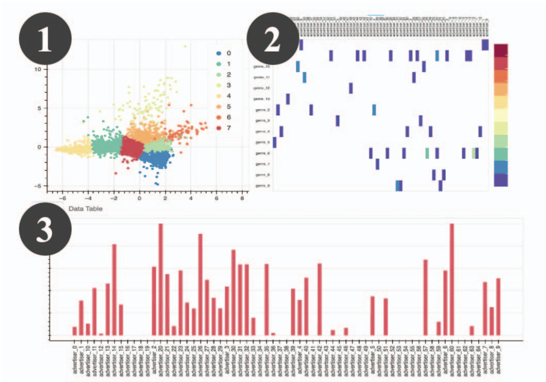
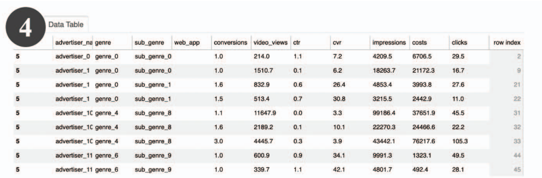


Fig. 1. Overall screen shot of the visualization system.



	advertiser_na	genre	sub_genre	web_app	conversions	video_views	ctr	cvr	impressions	costs	clicks	row index
5	advertiser_0	genre_0	sub_genre_0		1.0	214.0	1.1	7.2	4209.5	6706.5	29.5	2
5	advertiser_1	genre_0	sub_genre_0		1.0	1510.7	0.1	6.2	18263.7	21172.3	16.7	9
5	advertiser_1	genre_0	sub_genre_1		1.6	832.9	0.6	26.4	4853.4	3993.8	27.6	21
5	advertiser_1	genre_0	sub_genre_1		1.5	513.4	0.7	30.8	3215.5	2442.9	11.0	22
5	advertiser_1C	genre_4	sub_genre_8		1.1	11647.9	0.0	3.3	99186.4	37651.9	45.5	31
5	advertiser_1C	genre_4	sub_genre_8		1.6	2189.2	0.1	10.1	22270.3	24466.6	22.2	32
5	advertiser_1C	genre_4	sub_genre_8		3.0	4445.7	0.3	3.9	43442.1	76217.6	105.3	33
5	advertiser_11	genre_6	sub_genre_9		1.0	600.9	0.9	34.1	9991.3	1323.1	49.5	44
5	advertiser_11	genre_6	sub_genre_9		1.0	339.7	1.1	42.1	4801.7	492.4	28.1	45

Fig. 2. Data table displayed by switching tabs in Fig. 1(3).

In this study, we developed the visualization system by using Bokeh, a visualization library for Python. We used Python 3.9.6 and Bokeh 2.4.3. Fig. 1 shows the overview of the visualization system in this study. By switching tabs in the visualization screen at the bottom of Fig. 1, the data table can be displayed as shown in Fig. 2. The four visualizations (1) to (4) are described below.

1) The advertising data is visualized in the scatterplots after applying dimensionality reduction and clustering methods described in section 3.3. The advertisements which have similar abandonment behaviors belong to the same cluster. A unique color is assigned to each cluster in the scatterplots.
2) One of the clusters plotted in Visualization 1) is selected, and the click through rate of the advertisement data belonging to this cluster is visualized in the heatmap. In the visualization screen shown in Fig. 1, the vertical axis depicts the attribute of the advertisement genre and the horizontal axis depicts the attribute of the advertiser name. As shown in Fig. 3, red color is assigned to combinations of attributes with high click through rates, and purple color is assigned to combinations of attributes with low click through rates.

Fig. 3. Color map used in the heat map.

3) The results of visualizing the conversion rate for each advertiser are displayed as a bar graph. The selected data items are highlighted in the bar graph by selecting data items in the heatmap in Visualization 2).
4) The details of the data items displayed in Visualization 2) are displayed in the data table.

IV. Results

This section describes several results and discoveries obtained from the visualization. As an example, this paper presents the results of visualization and analysis of 224,355 anonymized records of video adcertisement dataset distributed by Yahoo! JAPAN Ads and 89,400 anonymized records of video advertisement data distributed by LINE Ads.

A. Results of visualization of video advertisements distributed by Yahoo! JAPAN Ads

We calculated the six variables R_1 to R_6 related to the video abandonment rate by using the method described in Section 3.2.1 with the anonymized data from the video ad data distributed by Yahoo! JAPAN Ads. Fig. 4 is the result of applying principal component analysis and clustering to the six values and then visualizing them in a scatterplot. Table II shows the results of calculating the eigenvectors of the first principal component (PC1) and the second principal component (PC2). PC1 has higher absolute values of R_1 to R_3, while PC2 has relatively higher absolute values of R_4 to R_6. Therefore, we found that PC1 is the abandonment rate in the entire video and PC2 is the abandonment rate in the second half. In the first principal component of Fig. 4, the further to the left, the advertisement data which have the higher abandonment rate is plotted. Further to the right, the advertisement data which has the lower abandonment rate is plotted. Fig. 5 shows the heatmap visualization of the click through rates of advertisements belonging to Cluster 1 which has the lowest abandonment rate from video advertisements out of the eight clusters visualized in the scatterplot in Fig. 4. The vertical axis depicts the genre of the video advertisement while the horizontal axis depicts the advertiser's name. This result shows that there exist differences in click through rates even among advertisements with low video ad abandonment rates. Specifically, advertisements with high click through rates correspond to those with "genre_8" and "advertiser_12" and those with "genre_4" and "advertiser_4".

Next, we selected cluster 3 which has a high abandonment rate of video advertisements. Fig. 6 shows the results visualized in the heatmap. The figure shows that there exist combinations of attributes with high click through rates, such as "genre_13" and "advertiser_48, even in the advertisement data with high advertisement withdrawal rates. It suggests that advertisements with such combinations of attributes have a high number of times in which abandonment occurred as a result of being clicked. Therefore, advertisements with these combinations of advertising attributes are connected to clicks and are effective advertising activities.

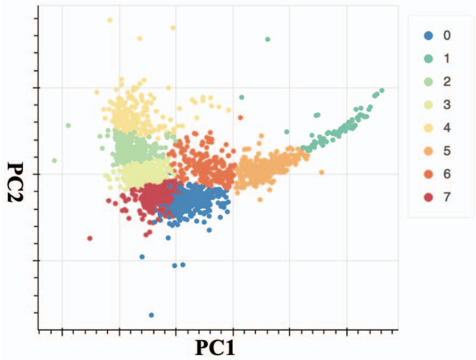


Fig. 4. The results visualized in scatter plots by applying principal component analysis and clustering by using dummy data in the same format as the advertising data delivered by Yahoo! JAPAN Ads.

TABLE II
EIGENVECTOR VALUES FROM PRINCIPAL COMPONENT ANALYSIS BY USING DUMMY DATA IN THE SAME FORMAT AS THE AD DATA DELIVERED BY YAHOO! JAPAN ADS.

	PC1	PC2
R1	-0.511	0.241
R2	-0.554	0.083
R3	-0.555	-0.008
R4	-0.287	-0.504
R5	-0.153	-0.613
R6	0.134	-0.553

Then, we selected the combination of ad attributes with high click through rates among the data items visualized in Fig. 6, and highlight the corresponding data items on the bar graph visualizing the conversion rate. From the visualization result shown in Fig. 7, we found that this combination of target attributes has a large click through rate and is also linked to conversions. This result demonstrates that users can comprehensively analyze advertising effectiveness in terms of the abandonment rate, click through rate, and conversion rate, by looking at the visualizations presented in this paper.

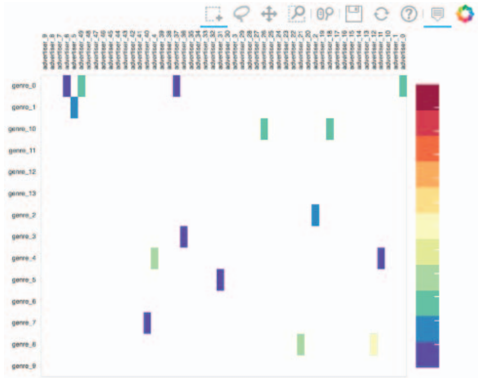


Fig. 5. The result of visualizing the click through rate of ad data which belongs to Cluster 3 in Fig. 4 by using a heatmap. (vertical axis: advertisement genre. horizontal axis: advertiser name.)

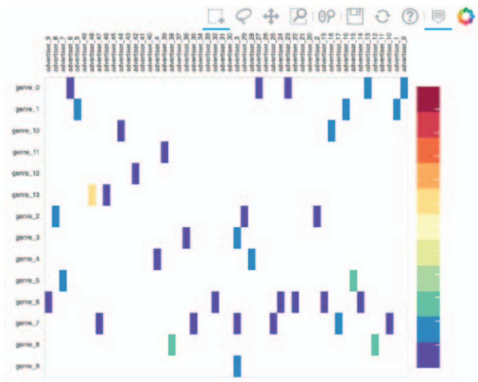


Fig. 6. The result of visualizing the click through rate of ad data which belongs to Cluster 3 in Fig. 4 by using a heat map. (vertical axis: advertisement genre, horizontal axis: advertiser name).

B. Visualization results for video advertisements distributed via LINE Ads

Fig. 9 shows the visualization results with anonymized data in the same format as the video ad data distributed by LINE Ads. Table III shows the eigenvectors of the principal component analysis. PC1 has high absolute values of R_1 to R_5 while PC2 has relatively high absolute values of R_6. In other words, PC1 corresponds to the abandonment rate in the entire video while PC2 corresponds to the abandonment rate in the latter half. Further to the left in Fig. 9, the lower abandonment rate from video advertisements is plotted in the advertisement data. Further to the right, more advertisements with high abandonment rates from video advertisements are plotted. Fig. 10 shows the results of a heatmap visualization of the click through rate for advertisements belonging to Cluster 4 which have the lowest abandonment rate. The vertical axis depicts

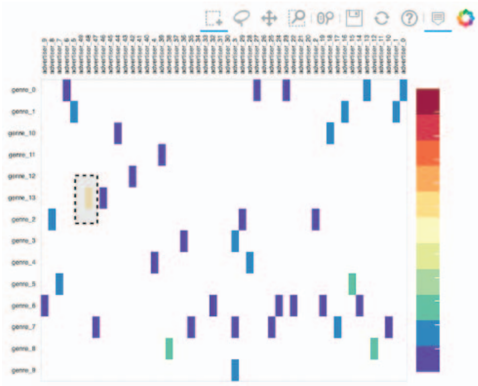


Fig. 7. Select the combination of attributes with a high click through rate among the data belonging to cluster 3 which has a high abandonment rate.

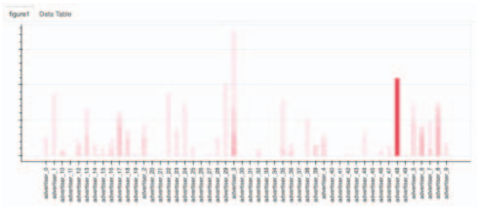


Fig. 8. The result of visualization highlighting the advertiser's conversion rate for the selected data in Fig. 7.

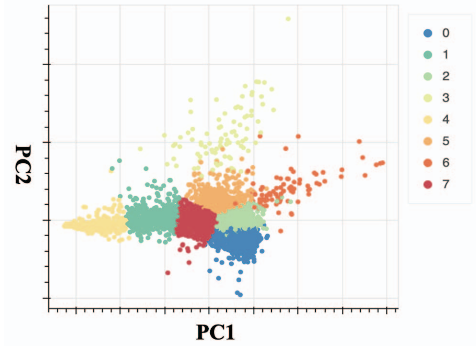


Fig. 9. The results visualized in scatter plots by applying principal component analysis and clustering by using dummy data in the same format as the advertising data delivered by LINE Ads.

TABLE III
EIGENVECTOR VALUES FROM PRINCIPAL COMPONENT ANALYSIS BY USING DUMMY DATA IN THE SAME FORMAT AS THE AD DATA DELIVERED BY LINE ADS.

	PC1	PC2
R1	0.408549	-0.428766
R2	0.475229	-0.294864
R3	0.495529	-0.114966
R4	0.402646	0.136094
R5	0.409860	0.477465
R6	0.177725	0.685203

the advertisement genre while the horizontal axis depicts the advertiser name. These visualization results show that the data distributed by LINE Ads are dominated by advertisements with relatively high click through rates.

Fig. 11 shows the visualization results after changing the attributes of the vertical and horizontal axes in the heatmap. The vertical axis depicts the attribute of the advertisement sub-genre while the horizontal axis depicts the attribute of the targeting type in Fig. 11. Table IV shows a description of the targeting type attributes. The visualization results show that the combination of advertisement attributes "sub_genre_4" and "Look Alike" has the highest click through rate. This result demonstrates that our visualization system enables users to discover the advertisement settings with the highest click through rate for each combination of the two attributes by changing the attributes on the vertical and horizontal axes of the heatmap.

V. CONCLUSION AND FUTURE WORK

We developed a visualization system for the evaluation of advertisements comprehensively by combining the abandonment rate, click through rate, and conversion rate of video advertisements. This paper introduced visualization results considering the combination of advertisement attributes. As a result, we found the following trends:

- There are video advertisements with high click through rates among video advertisements with high abandonment rates. Such video advertisements may have abandonment caused by a click.
- Different combinations of advertisement attributes among the platform may occur high click through rates.

We have three future issues. The first is to select the most appropriate dimensionality reduction and clustering method in this study. The visualization results vary greatly depending on which dimensionality reduction and clustering method are applied [7]. Therefore, it is necessary to evaluate whether the currently used method is appropriate. In the future, we would like to compare more diverse methods and seek the appropriate method for this study. The second is to improve the interactivity of the visualization tools. The current visualization tool does not allow the interactive selection of data items. We would like to enhance this tool to allow users to freely select data items so that users can archive more effective

TABLE IV
EXPLANATION OF TARGETING TYPE.

RT (Re Targeting)	Advertisements that display once-seen products, etc.
LA (Look Alike)	Advertisements that display items similar to those that consumers have seen or purchased in the past
AT (Audience Targeting)	Advertisements other than the above two

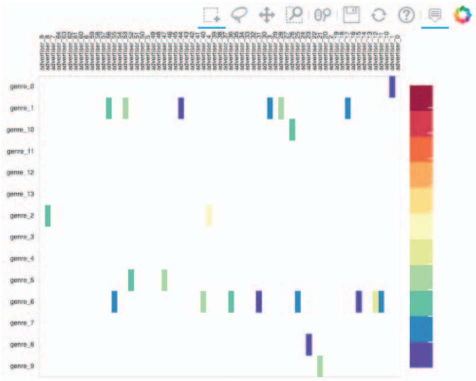


Fig. 10. The result of visualizing the click through rate of the advertisement data belonging to cluster 4 in Fig. 9 by using a heat map.(vertical axis: ad genres, horizontal axis: advertiser name)

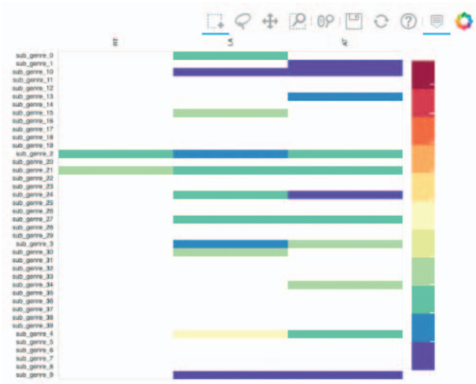


Fig. 11. The result of visualizing the click through rate of the advertisement data belonging to cluster 4 in Fig. 9 by using a heat map.(vertical axis: ad subgenres, horizontal axis: targeting types)

visualization analysis. The third is to apply parallel coordinate plots to visualization (3). Parallel coordinate plots will allow users to interpret the relationships among data items without losing information on multidimensional data, thus enabling more detailed data analysis.

References

[1] Dentsu: Advertising Expenditures in Japan in 2021 Internet Advertising Media Expenditures Detailed Analysis (2022), in Japanese, https://www.dentsu.co.jp/news/item-cms/2022007-0309.pdf, (referenced 2022-11-01).

[2] Javad Azimi, Ruofei Zhang, Yang Zhou, Vidhya Navalpakkam, Jianchang Mao, Xiaoli Fern, "Visual appearance of display ads and its effect on click through rate", In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pages 495–504, 2012.

[3] Junxuan Chen, Baigui Sun, Hao Li, Hongtao Lu, Xian-Sheng Hua, "Deep CTR Prediction in Display Advertising", In Proceedings of the 24th ACM international conference on Multimedia, pages 811–820, 2016.

[4] Muller Y.M. Cheung, Shuk Ying Ho, "Characteristics of Advertisements and Interactivity of Videos in Online Video Websites", SIGHCI 2017 Proceedings. 7, 2017.

[5] Zenichiro Ishikawa, Makito Oku, Takashi Kono, "Conversion Analysis and Forecasting Using Basic Advertisements data in Web Listing Advertisement", DEIM Forum, 2017, in Japanese.

[6] Shunsuke Kitada, Hitoshi Iyatomi, Yoshifumi Seki, "Conversion Prediction Using Multi-task Conditional Attention Networks to Support the Creation of Effective Ad Creatives", In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '19, pages 2069–2077, New York, NY, USA, 2019.

[7] Sylvain M. Lespinats, "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings", Computer Graphics Forum, pages 113-125, 2011.

[8] Shangsong Liu, Di Peng, Haotian Zhu, Xiaolin Wen, Xinyi Zhang, Zhenghao Zhou, Min Zhu, "MulUBA: multi-level visual analytics of user behaviors for improving online shopping advertising", Journal of Visualization, vol.24(6), pages 1287-1301, 2021.

[9] Yuka Matsumoto, Rikima Doumae, Akira Otuki, "Characteristic analysis of consumers using x-means clustering and principal component analysis", Japan Society of Information and Knowledge, pages 200-207, 2018, in Japanese.

[10] Shon Minjung, Lee Daeho, "Perceived Consumer Discomfort of Online Video Advertising: Analysis of Characteristics by Consumer Cluster", Journal of Scientific and Industrial Research, pages 528-536, 2021.

[11] Di Peng, Wei Tian, Min Zhu, Yukun Ren, Xiaojian Lin, Mingzhao Li, "TargetingVis: visual exploration and analysis of targeted advertising data", Journal of Visualization, vol.23(6), pages 1113-1127, 2020.

[12] Eric Zeng, Tadayoshi Kohno, Franziska Roesner, "What makes a "Bad" Ad? User Perceptions of Problematic Online Advertising", CHI Conference on Human Factors in Computing Systems, pages 1-24, 2021.

[13] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai, "Deep interest evolution network for click-through rate prediction", In Proceedings of the 33nd AAAI Conference on Artificial Intelligence, 2019.